ELSEVIER

Journal of Clinical Epidemiology 64 (2011) 366–379

Journal of Clinical Epidemiology

Critical appraisal of health literacy indices revealed variable underlying constructs, narrow content and psychometric weaknesses

Joanne E. Jordan[a,*], Richard H. Osborne[a,b], Rachelle Buchbinder[c,d]

[a]*Department of Medicine, Centre for Rheumatic Diseases, The Royal Melbourne Hospital, The University of Melbourne, Parkville, Victoria, Australia*
[b]*Public Health Research and Evaluation and Policy Cluster, School of Health and Social Development, Deakin University, Melbourne, Victoria, Australia*
[c]*Monash Department of Clinical Epidemiology at Cabrini Hospital, Monash University, Victoria, Australia*
[d]*Department of Epidemiology and Preventive Medicine, School of Public Health and Preventive Medicine, Monash University, Cabrini Medical Centre, Victoria, Australia*

Accepted 13 April 2010

Abstract

Objective: Health literacy refers to an individual's ability to seek, understand, and use health information. A range of indices exist that purport to measure health literacy across individuals and populations. This study aimed to review the development and content of existing indices and to critically appraise their properties.

Study Design and Setting: Using standardized search terms, published generic health literacy indices (1990–2008) were identified. Using a methodological framework, each was evaluated for purpose, validity (face, content, construct), reliability, responsiveness, feasibility, and generalizability.

Results: Nineteen instruments were evaluated. Three measurement approaches were identified: direct testing of individual abilities, self-report of abilities, and population-based proxy measures. Composition of underlying constructs and content varied widely across instruments, and none appeared to fully measure a person's ability to seek, understand, and use health information. The content was focused primarily on reading comprehension and numeracy; scoring categories were poorly defined and may not be mutually exclusive, and few indices had been assessed for reliability.

Conclusion: Health literacy is not consistently measured, making it difficult to interpret and compare health literacy at individual and population levels. Empirical evidence demonstrating validity and reliability of existing indices is required, and more comprehensive health literacy instruments need to be developed.

Keywords: Health literacy; Tools; Assessments; Psychometrics; Validity; Reliability

1. Introduction

As a result of health care changes toward a more patient-centered care approach [1,2], there is an expectation that patients take an increased role in decisions about their health. For patients to function effectively within this shifting paradigm, they require a basic set of skills to seek, understand, and use health information, a concept referred to as "health literacy" [3,4]. Suboptimal health literacy skills reduce the likelihood of maintaining good health, independent of other sociodemographic factors [5,6], and are associated with increased health care costs [7].

A range of indices have been developed to measure health literacy [8,9]. To provide credible information to inform clinical practice and health policies and programs, it is imperative that indices have sound psychometric properties. These include the following: validity—the extent to which an instrument measures what it purports to measure when properly administered [10]; reliability—the extent to which the obtained scores are free from measurement error [11]; feasibility—ease of administration; generalizability—use across and within fields; and in some circumstance, responsiveness—the ability of indices to detect change [12]. The purpose of this work was to review the literature on existing health literacy indices and perform a critical appraisal of their concept, content, and psychometric properties.

2. Method

Medline, PubMed, and PsychInfo databases were searched for publications of generic health literacy indices

* Corresponding author. Department of Medicine (RMH/WH), Centre for Rheumatic Diseases, Royal Melbourne Hospital, The University of Melbourne, Parkville, Victoria 3050, Australia. Tel.: +61-412-344-965; fax: +61-3-9569-2380.

E-mail address: joannejordan14@gmail.com (J.E. Jordan).

0895-4356/$ - see front matter
doi: 10.1016/j.jclinepi.2010.04.005

What is new?

Key finding

A key finding of this critical appraisal was that limited empirical evidence exists on the reliability and construct validity of health literacy measures. This raises uncertainty about the accuracy of data being produced in relation to health literacy levels at an individual and population level. Furthermore this research demonstrates that great variation exists across indices in terms of content across domains, methods of item selection and scoring and how individuals are categorised or classified in terms of health literacy levels. The variation and weak psychometric data makes it difficult to compare or pool results across studies and hinders the establishment of clear benchmarks for policy and program development aimed at addressing suboptimal health literacy.

What this adds to what is known?

This appraisal, using a structured methodological framework, provides new and synthesised information for researchers and clinicians on the strengths and limitations of current indices of health literacy. Although previous literature acknowledges the narrow approach taken to the measurement of health literacy, the systematic presentation of the validity, reliability and applicability of these instruments provides more detail relating to psychometric properties of existing instruments.

What is the implication, what should change now?

This review will support clinicians, researchers and policymakers to qualify the findings and recommendations derived from the use of health literacy indices. Further research is required to obtain empirical evidence across different populations and settings of the construct validity and reliability of existing measures. New measures which incorporate broader constructs of health literacy are required and are needed to advance this field.

between 1990 and 2008 inclusive, using the following search terms: health, literacy, health literacy, instrument, tool, and assessment. The search was limited to publications in English language. Instruments that were not generic, that is, specific to particular fields or groups, were excluded. Additional instruments were identified through manual searching of the references of relevant published studies and by consulting experts in the field.

All health literacy indices were evaluated for purpose, face validity, content validity, construct validity, reliability, responsiveness, feasibility, and generalizability, using a methodological framework developed for the evaluation of health assessment indices [13]. Categories within the framework were adapted where necessary to make the appraisal relevant for health literacy indices (Tables 1 and 2). Two raters (J.E.J. and R.B.) critically appraised all instruments independently, and disagreements were resolved by consensus or discussion with the third author (R.H.O.).

3. Results

3.1. Questionnaires

Nineteen health literacy instruments met the prespecified selection criteria and were included in the appraisal (Fig. 1) (Tables 1, A1, and A2). There were 12 original instruments [8,9,14–23] and seven derivatives that were short-form versions or adaptations of original instruments [24–30]. There were also Spanish versions of two of the instruments [9,16] and a Chinese version of a short-form instrument [31]. As these were direct translations of original instruments, they were excluded.

There were three main approaches for measuring health literacy: direct testing of an individual's abilities (Table 1), elicitation of self-report of abilities (Table A1 [available on the journal's Web site at www.jclinepi.com]), and proxy measures of health literacy in the population (Table A2 [available on the journal's Web site at www.jclinepi.com]).

3.1.1. Instruments that directly test an individual's abilities

3.1.1.1. Rapid Estimate of Adult Literacy in Medicine and derivatives. The Rapid Estimate of Adult Literacy in Medicine (REALM) was developed in the United States to assist physicians to identify adults with limited reading skills in the primary care setting (Table 1) [8]. It was modeled on the Wide Range Achievement Test (WRAT), a standardized literacy test that assesses the pronunciation of a list of words that the respondent reads aloud [8]. The 125 words, chosen from materials used in primary care, are arranged in three columns according to the number of syllables in ascending order of difficulty. A point is allocated for each correctly pronounced word. A shortened version, the REALM-S, comprising 66 items, was developed to reduce administration time and enhance uptake within the clinical setting [24]. It takes 1–2 minutes to complete. The raw score for the REALM-S (0–66) is converted to a US school grade estimate of reading ability (Table 1).

Other derivatives include the REALM-shortened version (REALM-R) (eight items) [26] and REALM-Short Form (REALM-SF) (seven items) [27]. The Rapid Estimate of Adolescent Literacy in Medicine (REALM-Teen), another derivative, is tailored to screen youth in grade 6–12 (ages 10–19 years) for below-grade reading in the health care setting [25]. Words for the REALM-Teen were selected

Table 1
Selected results of characteristics of health literacy indices that directly test individual abilities

Name of instrument	REALM [8]	REALM-S [24]	REALM-Teen [25]	SAHLSA [15]	MART [14]	TOFHLA [9]	S-TOFHLA [30]	HHLT [29]	NVS [16]
Country of origin	USA	USA	USA	USA	USA	USA	USA	Israel	USA
Year of publication	1991	1993	2006	2006	1997	1995	1999	2007	2005
Stated expertise of developers	Not stated	Not stated	Panel of doctors, nurses, social workers, psychologists, and educators	Experience working with Spanish-speaking patients in educational and medical settings	Not specified	Literacy expert	Not stated	Public health experts	Panel of health literacy experts
1. Purpose									
a. Stated purpose and population	Identify patients with limited reading skills and estimate patient reading levels in primary care setting	Identify patients with low reading levels in primary care settings	Screen youth in grades 6–12 for below-grade reading	Develop an easy-to-use health literacy test for Spanish speakers in health care settings	To identify illiterate patients of high school age or older in the general community	Understand and measure functional health literacy in patients in health care setting	Measure patients' ability to read and understand health-related materials in the health care system	Assess health literacy in Hebrew patients in Israeli health care system	Screen for limited literacy in patients in primary health care settings
2. Method of development									
a. How was the instrument developed (or shortened if more applicable)	Words in instrument chosen from patient education materials and patient intake forms used in university-based primary care clinics. Method of selection not stated.	Shortened using psychometric estimate of item difficulty and discrimination and frequency of retained words in written material given to patients	116 words were selected from American Academy of Paediatrics' adolescent patient education pamphlets. The list was piloted tested on 200 students in grades 6–12 and retention of items based on psychometric estimate of item difficulty, item discrimination, and the panel's judgement.	Items from REALM translated to Spanish. For the additional comprehension section, words were selected by an expert panel both fluent in English and Spanish using the Delphi process	Based on the Wide Range Achievement Test. Words selected from prescription labels and from a medical dictionary. The 42 words were chosen to have a corresponding level of difficulty based on words contained in the WRAT reading test.	Developed from sample of hospital texts, including education materials, diagnostic tests, prescription labels, registration forms by literacy expert	Item selection based on previous data from large-scale study that used the TOFHLA. Numeracy items selected on perceived importance and frequency of task, proportion of items answered incorrectly from prior study and ease of administration	Based on S-TOFHLA. Three numeracy items directly translated from S-TOFHLA, 1 item changed to resemble Israeli style of arranging appointments. Reading comprehension passages developed by authors	Five health-related scenarios were developed by a panel of health literacy experts. Scenarios refined through consultation and scenario selected that was found to have strongest psychometric properties

3. Description of instrument									
a. Domain	Single domain—125 items.	Single domain—66 items.	Single domain—66 items	Single domain—50 items	Single domain—42 items	Two domains: (1) reading comprehension—50 items; (2) numeracy—17 items	Two domains: (1) reading comprehension—36 items; (2) numeracy—4 items	Two domains: (1) reading comprehension—8 items; (2) numeracy—4 items	Single domain—6 items
4. How is it administered	Interview administered	Interview administered	Interview administered	Interview administered	Interview administered	Numeracy section interview administered	Numeracy section interview administered	Interview administered	Interview administered
5. Special requirements for administration	Knowledge of correct pronunciation of words	Knowledge of correct pronunciation of words	Knowledge of correct pronunciation of words	Knowledge of correct pronunciation of words and provide cue cards	Knowledge of correct pronunciation of words	Provide cue cards and verbally administer numeracy questions	Provide cue cards and verbally administer numeracy questions	None specified	Verbally administer questions
6. Estimated time duration	3−5min	1−2min	2−3min	3−6min	3−5min	Upto 22min	Less than 10min	Not stated	Average 2.9min
7. Scoring									
a. How is it scored	Single score: 0−125	Single score: 0−66	Single score: 0−66	Single score: 0−50	Single score: 0−42	Combined weighted score: 0−100	Combined weighted score: 0−100	Combined score from 0 to 12	Single score: 0−6
b. Scoring categories	School grade estimate: 0−78 below third grade; 79−103 for fourth to sixth grade; 104−114 for seventh to eighth grade; 115+ for high school	School grade estimate: 0−18 for third grade and below; 19−44 for fourth to sixth grade; 45−60 for seventh to eighth grade; 61−66 for ninth grade and higher	School grade estimate: 0−37 below third grade; 38−47 for fourth to fifth grade; 48−58 for sixth to seventh grade; 59−62 for eighth to ninth grade; 63−66 for 10th grade and higher	2 categories—<37: inadequate health literacy; 38−50: adequate health literacy	School grade levels but categories are not specified	0−59: inadequate health literacy; 60−74: marginal health literacy; 75−100 adequate health literacy	0−53: inadequate functional health literacy; 54−66: marginal health literacy; 67−100: adequate health literacy	0−2: low health literacy; 3−10: marginal health literacy: 11−12 high health literacy	0−1: high likelihood of marginal/inadequate literacy; 2−3: possibility of marginal/inadequate literacy; 4−6: adequate literacy

Abbreviations: REALM, Rapid Estimate of Adult Literacy in Medicine; REALM-S, Rapid Estimate of Adult Literacy in Medicine shortened version; REALM-Teen, Rapid Estimate of Adolescent Literacy in Medicine; SAHLSA, Short Assessment of Health Literacy for Spanish-speaking Adults; MART, Medical Achievement Reading Test; TOFHLA, Test of Functional Health Literacy in Adults; S-TOFHLA, Test of Functional Health Literacy in Adults short version; HHLT, Hebrew Health Literacy Test; NVS, Newest Vital Sign.

Table 2
Selected results of critical appraisal of indices that directly test patient abilities

Name of instrument	REALM [8]	REALM-S [24]	REALM-Teen [25]	SAHLSA [15]
2. Is the instrument based on an underlying conceptual framework?	No	No	No	Not specified
3. Content validity				
a. Were intended domains (i.e., relevant areas to be included and excluded) clearly stated?	No	No	No	No
b. Are all relevant components of each domain included?	Yes	Yes	Yes	Yes
4. Face validity				
a. On the face of it does it describe the intended purpose	Yes	Yes	Yes	Yes
b. Are the definitions of each category clearly specified?	Partial	Partial	Partial	No
c. Are these definitions acceptable?	Partial—provides a school grade estimate for reading ability with descriptions for whether individual may/may not need low literacy materials	Partial—no clear differentiation between 4–6 and 7–8 grade range estimates in terms of what materials will be suitable	Partial—no accompanying definitions with school grade range estimates	No
d. Are the methods for determining the presence and/or absence of criteria described and acceptable?	No. Criteria not described and correct pronunciation of words will vary amongst population groups	Partial, dictionary pronunciation is the scoring standard for ruling whether words are pronounced correctly	Partial, dictionary pronunciation is the basis of scoring standard	Partial, didn't fully detail how different idiomatic expressions across Spanish-speaking countries were adapted
e. Do the scoring categories sufficiently discriminate considering the stated purpose?	Partial—hard to distinguish between grade equivalents 4–6 and 7–8 in terms of what health information materials would be used	Partial. The categories do not provide clear directions in terms of which groups categorically will need low literacy materials	Partial. Provides a school grade estimate but there is no corresponding description to assist what materials an individual may need.	Partial—identifies between inadequate and adequate health literacy but no accompanying descriptions
5. Construct validity				
a. Does the instrument perform in expected ways when compared with other health literacy/literacy indices?	Yes, high correlation with standardized general reading tests (1) SORT-R—Pearson correlation coefficient = 0.95; (2) PIAT-R—Pearson correlation coefficient = 0.94	Yes, high correlation with standardized general reading tests: (1) SORT-R—Pearson correlations $r = 0.96$; (2) PIAT-R—Pearson correlation $r = 0.97$, and (3) WRAT—Pearson correlation $r = 0.88$. Moderately high correlation with (1) TOFHLA—Spearman rank correlation = 0.74; (2) Basic Skills Assessment Initial Test (UK)—Pearson correlation coefficient = 0.70; (3) REALM-R—Spearman rank correlation = 0.72. Poor correlation with (1) TORCH (Aus)—Pearson correlation coefficient = 0.36	Yes high correlation with (1) WRAT-3—Pearson correlation $r = 0.83$ and (2) SORT-R—Pearson correlation $r = 0.93$	Partial, correlated with TOFHLA (Spanish version)—Pearson correlation $r = 0.65$. Difference expected given no numeracy component for SAHLSA. Confirmatory factor analysis model: RMSEA = 0.04 and CFI = 0.85
6. Has sensitivity to change been demonstrated?	No	No	No	No
7. Reliability				
a. Has the reliability been measured?	Yes, test-retest reliability coefficient 0.98	Yes, test-retest reliability coefficient 0.99	Yes, test-retest coefficient 0.98	Yes, test-retest reliability coefficient 0.86
b. Interrater reliability	0.99	Not stated	Not stated	Not stated
c. Intrarater reliability	N/A	N/A	Not stated	Not stated
d. Internal consistency	Not stated	Not stated	Yes, Coefficient alpha 0.94	
7. Feasibility				
a. Is it simple to understand?	Yes	Yes	Yes	Yes
b. Is it easy to perform and administer?	Partial, knowledge of correct pronunciation of words required.	Partial, knowledge of correct pronunciation of words required.	Partial, knowledge of correct pronunciation of words required	Partial, uncertainty in differentiating between different Spanish expressions

Abbreviations: SORT-R, Slosson Oral Reading Test; PIAT-R, Peabody Individual Achievement Test Revised; WRAT, Wide Range Achievement Test; TORCH, Test of Reading Comprehension; RMSEA, root mean square error of approximation; CFI, comparative fit index; HHLT, Hebrew Health Literacy Test.

MART [14]	TOFHLA [9]	S-TOFHLA [30]	HHLT [29]	NVS [16]
No	Functional health literacy, defined as reading, writing, and numeracy skills	Functional health literacy	Functional health literacy	Not specified
No	No	No	Yes	Partial, justification for inclusion of numeracy
Partial, illiteracy defined as unable to read and write but only look at reading ability.	Partial, writing ability (stated as component of functional health literacy) not tested	Partial, does not include testing of writing abilities	Partial, writing ability not tested.	Partial—overemphasis on numeracy items.
Partial	Partial	Partial	Partial	Partial
No	No	No	No	No
No	No	No	No	No
Partial, correct word pronunciation not specified	Yes	Yes	Yes	Yes
Unable to be evaluated. Based on WRAT scoring categories but scoring levels and categories not outlined	Partial—categorizes into inadequate, marginal, and adequate functional health literacy but no generalizable descriptions	Partial—categorize into inadequate, marginal, and adequate functional health literacy but no generalizable descriptions	Partial—categories differ from that of S-TOFHLA and use low/marginal/high health literacy categories but no accompanying definitions	Partial—there seems to be a bit of overlap for scores 0–1 and 2–3 in terms of both having descriptions of inadequate/marginal literacy. No definitions provided
No. No comparisons conducted.	Partial. Correlated with (1) WRAT—Spearman rank correlation = 0.74 and (2) REALM—Spearman rank correlation = 0.84. The latter correlation is possibly higher than expected given that the TOFHLA is measuring comprehension and numeracy as well as reading ability. Poor correlation with (1) TORCH (Aus)—Pearson correlation coefficient = 0.39 and (2) NVS Spearman rank correlation = 0.49	Yes, reading comprehension section correlates highly with REALM—Spearman correlation coefficient = 0.81 although low correlation for the numeracy section with REALM—Spearman correlation coefficient = 0.61 which is to be expected as measuring different constructs	Partial. When results of HHLT compared with proxy measure of patient self-reported reading ability. Pearson correlation $r = 0.67$	No, low correlation with (1) TOFHLA that measures similar constructs—Pearson correlation $r = 0.59$; (2) S-TOFHLA Pearson correlation coefficient $r = 0.6$; (3) REALM Pearson correlation coefficient $r = 0.41$ (primary care clinic) and Spearman rank correlation = 0.24 (Australian population study).
No	No	No	No	No
No	No	No	No	No
Not stated	N/A	N/A	N/A	N/A
Not stated	Not stated	Not stated	Not stated	Not stated
Yes, Coefficient alpha 0.98	Yes, Coefficient alpha 0.98	Yes, Coefficient alpha 0.68 (numeracy) and 0.97 (reading comprehension)	Yes, Coefficient alpha 0.98	Yes, Coefficient alpha 0.76
Yes	Yes	Yes	Yes	Yes
Partial, knowledge of correct pronunciation required.	Partial (length)	Yes	Yes	Yes

by an expert panel from adolescent patient education pamphlets, and it has more school grade categories than the REALM (Table 1).

3.1.1.2. Short Assessment of Health Literacy for Spanish-speaking Adults. The Short Assessment of Health Literacy for Spanish-speaking Adults (SAHLSA) was based on the REALM-S, but it also includes a comprehension test (Table 1) [15]. Respondents read aloud a list of 50 medical terms and also have to nominate a word of similar meaning from a choice of two other words. A point is allocated for each correct answer, and respondents are classified as having either inadequate or adequate health literacy [15].

3.1.1.3. The Medical Achievement Reading Test. Also modeled on the WRAT, the Medical Achievement Reading Test (MART) was designed to identify illiterate patients in a nonthreatening manner [14]. It consists of 42 words, selected from prescription labels and medical dictionary and chosen to reflect the same difficulty level as that of the items in the WRAT. For respondents to feel less intimidated when undergoing testing, the MART uses small print and a glossy cover that creates a glare and makes reading the words difficult to provide individuals with excuses for not being able to read the words. Similar to the WRAT, raw scores are converted to school grade levels; however, scoring and grade-level placement were not provided in the published literature [14,32].

3.1.1.4. Test of Functional Health Literacy in Adults and derivatives. The Test of Functional Health Literacy in Adults (TOFHLA) was developed in the United States to measure ''functional health literacy,'' defined as assessing reading, writing, and numeracy skills, in relation to health care (Table 1) [9]. Items were chosen from hospital texts by a literacy expert. A self-administered reading comprehension component includes three passages of texts and contains 50 items. It uses the modified Cloze procedure, where every fifth to seventh word in a passage is omitted, and the respondent selects a response from four options [33]. An interviewer-administered numeracy component includes 17 items, where individuals are presented with cues and respond to questions. Scores for the numeracy domain are transformed to a score out of 50 and added to the score from the reading comprehension domain (total: 0–100).

The raw score for the TOFHLA is converted to one of three categories: inadequate, marginal, or adequate health literacy. These categories were derived from an interpretation of the scores of 2,659 predominantly indigent and minority patients presenting for acute care at two urban hospitals in Atlanta and California, although the method is not explained [34].

A short version, S-TOFHLA, includes 36 reading comprehension and four numeracy items and are also converted to three categories although the score cutoffs are different (Table 1) [30]. Another derivative instrument, the Hebrew Health Literacy Test (HHLT) [29], retains the same structure as the S-TOHFLA, but the items have been adapted to be relevant to the Israeli health care system and is administered in Hebrew. The reading comprehension component has eight items, and the numeracy component has four items. Individuals are classified as having either low, marginal, or high health literacy.

3.1.1.5. Newest Vital Sign. The Newest Vital Sign (NVS) was developed by a panel of health literacy experts to screen for limited literacy in the primary health care setting [16]. It tests reading, interpretation, and numeracy skills. An interviewer asks six questions relating to information contained in a nutritional label from an ice cream container. A point is given for each correct answer. Scores are categorized as high likelihood of marginal/inadequate literacy, possibility of marginal/inadequate literacy, and adequate literacy (Table 1).

3.1.2. Questionnaires that elicit self-report of abilities

3.1.2.1. Set of brief screening questions and derivative. The set of brief screening questions (SBSQ) was developed to detect inadequate or marginal health literacy in a clinical setting (Table A1 [available on the journal's Web site at www.jclinepi.com]) [17]. The content of the 16 questions was based on five domains identified in a qualitative study of patients with limited health literacy [35]. Respondents are asked to report how frequently they experience problems reading or understanding health information or how confident they are performing these tasks using a 5-point Likert scale (Table A1 [available on the journal's Web site at www.jclinepi.com]). Based on the comparison with the S-TOFHLA, three questions were selected for the final instrument based on their greater sensitivity and specificity for detecting individuals with inadequate health literacy (Table A3).

A derivative of the SBSQ, the Single-Item Literacy Scale (SILS) asks respondents to report their need for help to read or understand printed health information [28] (Table A1 [available on the journal's Web site at www.jclinepi.com]). A cutoff score $\geqslant 2$ indicates difficulty with reading printed health-related information [28].

3.1.2.2. Scales for measuring functional, communicative, and critical health literacy. Developed in Japan by clinicians and researchers, the functional, communicative, and critical health literacy (FCCHL) measures three different levels of health literacy: functional, communicative, and critical health literacy [19]. It was developed specifically for patients with type 2 diabetes but has subsequently been adapted for other populations [36]. It is based on a theoretical model developed by Nutbeam [37]—see Box 1 for definitions. Respondents rate their functional (five items), communicative (five items) and critical (four items) health literacy abilities using 4-point scales [19]. A numerical score is obtained for each scale by averaging the item

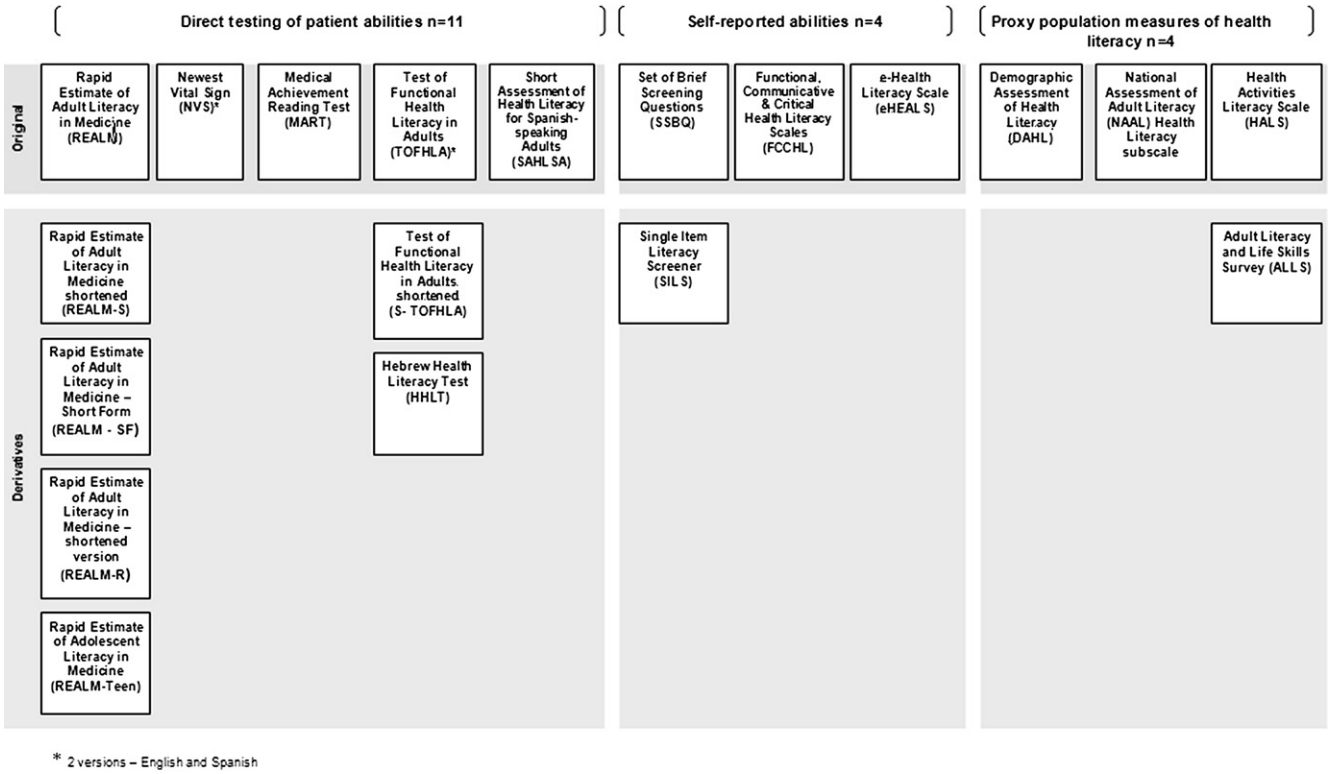

Fig. 1. Nineteen health literacy instruments (original and derivatives) included in the appraisal. Shortened or modified indices that were based on an existing assessment are referred to as ''derivative'' in this study. Instruments were also categorized based on the approach used to assess health literacy.

scores (Table A1 [available on the journal's Web site at www.jclinepi.com]).

3.1.2.3. eHealth Literacy Scale. The eHealth Literacy Scale (eHEALS) was developed in Canada and measures an individual's ability to use electronic health information for health care problems or issues in the broader population [18]. It is based on a model of eHealth literacy derived from social cognitive and self-efficacy theory (Table A3) [38]. Respondents are asked to indicate how strongly they agree or disagree with eight statements concerning their knowledge, confidence, and perceived skills relating to the use of electronic health information. It is unclear from the published literature how the eHEALS is scored and interpreted.

3.1.3. Proxy measures of health literacy in the population

3.1.3.1. The demographic assessment of health literacy. The demographic assessment of health literacy (DAHL) is a proxy indicator for health literacy based on the sociodemographic variables of age, sex, race/ethnicity, and years of schooling (Table A2 [available on the journal's Web site at www.jclinepi.com]) [20]. It was derived from a large US population-based study (individuals older than 65 years) based on the S-TOFHLA. The development of the algorithm used associations between S-TOFHLA and the four demographic variables to generate equations (scoring weights), which are then applied to surveys to produce an imputed health literacy score and estimate the presence of limited health literacy at a population level.

3.1.3.2. National Assessment of Adult Literacy. The 2003 National Assessment of Adult Literacy (NAAL), a national literacy survey undertaken in the United States, included 28 items to measure health literacy in the population (Table A2 [available on the journal's Web site at www.jclinepi.com]) [21]. These items address three health areas—clinical (three items), prevention (14 items), and navigation of the health care system (11 items)—and were developed to fit into one of the NAAL's prose, document, and quantitative scales (see Box 2 for definitions) [21]. Items are not publicly available [21,39]. Each health literacy item is mapped to the relevant prose, document, or quantitative scale, and categorized as either below basic, basic, intermediate, or proficient (Table A2 [available on the journal's Web site at www.jclinepi.com]) [21].

3.1.3.3. Health Activities Literacy Scale. The Health Activities Literacy Scale (HALS) was also developed in the United States to measure health literacy skills among adults in the population [22]. It consists of 191 health-related items, taken from large-scale adult literacy surveys conducted before 2003 and coded to represent health activities across five contexts or domains (Table A2 [available on the journal's Web site at www.jclinepi.com]). All items relate to prose, document, or quantitative literacy. The individual

item properties are unknown [40], and scoring properties for the HALS have not been specified [23].

3.1.3.4. Adult Literacy and Life Skills Survey. The Adult Literacy and Life Skills Survey (ALLS) includes a domain that measures health literacy in adult populations and has been administered in Canada and Australia [23,41]. It provides information on knowledge and skills across four areas: prose literacy, document literacy, numeracy, and problem solving. ALLS items appear to be a derivative of the HALS [23]. Proficiency of health literacy is measured on a 0–500 scale and grouped into five skill levels (Table A2 [available on the journal's Web site at www.jclinepi.com]) [41]. Items of this survey are also unavailable.

3.2. Critical appraisal of psychometric properties

A summary of the psychometric attributes for each instrument is outlined in Tables 2, A4, and A5 and discussed in the following sections.

3.2.1. Content validity

Content validity evolves out of the planning and construction of an instrument and looks at the extent to which the items making up the instrument reflect the concept that is intended to be measured [42]. The underlying conceptual framework of health literacy varied across instruments. The TOFHLA and its derivatives were based on a refined definition of functional health literacy (Table 2) [9]; the content of questions in the SBSQ was based on five domains identified in a previous qualitative study [17,35]; and both FCCHL and eHEALS were based on specified theoretical models of health literacy and eHealth literacy, respectively (Table A4 [available on the journal's Web site at www.jclinepi.com]) [37,38]. No clear underlying conceptual framework of health literacy was specified for the REALM, SAHLSA, NVS, or MART. The health literacy items selected for the NAAL were based on the definition of health literacy adopted in the US policy document "Healthy People 2010" (Table A5 [available on the journal's Web site at www.jclinepi.com]). As acknowledged by the authors, although this definition implies broader skills, such as knowledge, items were limited to prose, document, and quantitative literacy in keeping with the overall literacy focus [21].

For most of the indices, there was a good description of how items were generated. The item content for the REALM and TOFHLA were developed from commonly used patient materials in the health care setting, whereas the NVS used a generic ice cream nutritional label that was chosen from four other scenarios on the basis of having the strongest psychometric properties [16].

Based on the underlying constructs or definitions specified by the authors, some measures did not cover all relevant intended content. For example, the construct of the TOFHLA is stated to be reading, writing, and numeracy skills in relation to health care, yet it does not include writing abilities. Similarly, although the SBSQ was based on five domains, the final SBSQ consists of only three items, and the eight items of the eHEALS do not appear to fully represent the six different types of literacy it purports to measure. Table A3 (available on the journal's Web site at www.jclinepi.com) presents the constructs and items for SBSQ and eHEALS.

For population proxy indices, the DAHL presents as having good content validity, given it uses a credible range of demographic variables to derive a measure of health literacy. However, the weights across variables were derived from one sample of people aged 65 years or older, and this may not be generalizable to other ages and/or other settings. The content validity was unable to be evaluated for the other population surveys, as items are not publicly available [39].

For derivative instruments, how items were retained or excluded was well described and was based on psychometric estimates of item difficulty or statistical analyses.

3.2.2. Face validity

Face validity refers to whether an instrument appears to test what it is supposed to and that it is a plausible method for doing so [42]. Although the formats of each instrument appear plausible for their stated purpose, the face validity of most indices was questionable because of the difficulty in interpreting their scores. Most instruments collapse numerical scores into one of several categories but do not provide specific definitions for those categories. For some instruments, it was not possible to determine whether the categories were independent or mutually exclusive.

The REALM and its derivatives provide the most detailed definitions of their scoring categories. Categories correspond to school grade equivalents (Table 1) and have accompanying descriptions as to how simple the material or instructions should be to cater to different levels [8]. Conversely, the TOFHLA and its derivatives classify an individual as having either inadequate, marginal, or adequate health literacy, but provide no operational definitions of these categories. Instead, they identify tasks contained within the TOFHLA that individuals with suboptimal health literacy might have difficulty with. For example, patients with inadequate functional health literacy are expected to often misread medication dosing instructions, whereas those with marginal literacy are expected to have difficulty comprehending the Medicaid passage [34]. These examples are limited and may not be applicable to all individuals in these categories.

The SAHLSA used the TOFHLA to determine the cutoff scores between inadequate and adequate health literacy, but no descriptions for these categories are provided. The NVS scoring categories refer to the likelihood of inadequate or marginal literacy, but without clear accompanying definitions, these classifications appear to overlap. The FCCHL produces numerical scores for each level of health literacy.

No definitions are provided for scores; hence, it is unclear what constitutes appropriate health literacy levels.

For population surveys, scoring categories relate to ''proficiency'' of skills. Although definitions are outlined, these tend to be generic in terms of whether individuals have skills to be able to carry out tasks in everyday life rather than specific details relating to health-related activities.

The face validity of the HALS and MART could not be assessed. The HALS scoring algorithm is yet to be defined, and although the MART scoring system is based on the WRAT, specific details of categories and classifications were not outlined.

3.2.3. Construct validity

Construct validity refers to the degree to which an instrument measures what it is supposed to measure and can be assessed in various ways through factor analysis, hypothesis testing, and examining associations an instrument has with existing measures, where high correlation between tests (>0.7) is indicative of a similar construct being measured [42,43].

The construct validity of health literacy indices that directly test an individual's abilities has predominantly been assessed by comparing them with standardized reading tests and other health literacy instruments [8,9,15,16,24–27,30,44–48] (Table 2). Both the REALM-S and TOFHLA have demonstrated strong correlation with the WRAT in adult populations in the United States—Pearson correlation coefficient = 0.88 and Spearman rank correlation = 0.74, respectively [9,24]. However, lower correlations with the WRAT were found for both measures in an adolescent population in the United States (Pearson correlation coefficients—REALM: 0.74 and TOFHLA [reading comprehension domain]: 0.6) [45]. The same measures have also been shown to weakly correlate with the Test of Reading Comprehension, an Australian reading assessment used with students in years 3–10 (Pearson correlation coefficients—TOFHLA: 0.39 and REALM: 0.36) [44,49]. The REALM has also correlated highly with the Slosson Oral Reading Test Revised (SORT-R); Peabody Achievement Test Revised; and the Basic Skills Assessment Initial Test, a measure of general literacy used in the United Kingdom [46] (Table 2). Derivatives of the REALM have also demonstrated high correlation with the WRAT [25,26]. The MART has not been compared with other literacy tests, and the SAHLSA used factor analysis (see later) to demonstrate construct validity.

Comparison of ''direct testing'' health literacy indices with one another has yielded variable results across studies. Both the TOFHLA and the reading comprehension domain of the S-TOFHLA have been shown to correlate highly with the REALM (Spearman rank correlation: 0.84 and 0.81, respectively) in the original studies [9,30]; however, subsequent studies have reported low to moderate correlations between the REALM and TOFHLA [44,45,47]. For example, the correlation between the reading comprehension section of the TOFHLA and REALM was moderate (Pearson correlation coefficient: 0.6) in an adolescent population study in the United States [45] and low in a study of patients attending a community-based rheumatology practice in Australia (Pearson correlation coefficient: 0.30) [44] and another Australian population-based study (Spearman rank correlation: 0.31) [47].

The NVS has been shown to correlate moderately with the TOFHLA in the US primary care study (Pearson correlation coefficient: 0.59) [16] and an Australian population-based study (Spearman rank correlation: 0.49) [47]. The NVS has also been shown to correlate moderately with the S-TOFHLA (Pearson correlation coefficient: 0.61) [48]. In contrast, the NVS has demonstrated low correlation with the REALM (Pearson correlation coefficient of 0.41 and Spearman rank correlation of 0.24, respectively) [47,48].

The SBSQ, SILS, and DAHL used the S-TOFHLA as the standard to assess the specificity and sensitivity of screening questions or model in the case of the DAHL (Tables A4 and A5 [available on the journal's Web site at www.jclinepi.com]). For the three selected questions of the SBSQ, the optimum sensitivity and specificity appeared to be for the response ''sometimes/somewhat'' [17]. Subsequent testing of these questions with the REALM and S-TOFHLA suggests one possible question—''confidence with filling out forms'' for detecting inadequate health literacy (area under the receiver operating characteristic curve = 0.74 [95% confidence interval (CI): 0.69–0.79] and 0.84 [95% CI: 0.79–0.89], respectively) [50]. For the SILS, a score $\geqslant 2$ had 54% sensitivity and 83% specificity for detecting inadequate health literacy [28]. A cutoff score less than 62 in the DAHL resulted in 79% cases positively classified with a low sensitivity of 58% and an acceptable specificity of 84% [20].

Confirmatory factor analysis was undertaken to demonstrate construct validity for the REALM-SF and SAHLSA. This analysis identifies whether distinct underlying constructs or factors are identifiable in the data. Standardized fit statistics, for example, root mean square error of approximation (RMSEA), standardized root mean square residual (SRMR), and comparative fit index (CFI), provide evidence of whether distinct hypothesized constructs are present (see Ref. [51]). Both the REALM-SF (SRMR = 0.01, CFI = 0.97) and SAHLSA (RMSEA = 0.04, CFI = 0.85) have been shown to have good model fit (Table 2) [15,27].

For both the FCCHL and HHLT, demonstration of construct validity was limited, because other health literacy measures have not been translated into relevant languages. For the HHLT, scores were moderately correlated with patient self-reported reading ability (Pearson correlation coefficient: 0.67) [29].

3.2.4. Reliability

Reliability refers to the extent to which an instrument is consistent and free from error [42]. There are several approaches to reliability: establishing an instrument is capable of obtaining consistent results on separate occasions from

individuals in a stable environment (test–retest reliability) [52], assessing consistency of the results obtained by the same rater (intrarater reliability) and the agreement between the results of different raters (interrater reliability) [53]. Reliability coefficients generally range between 0.0 and 1.0, where 0.0 indicates that all measurement variation is attributable to error, and 1.0 indicates that the measurement has no error.

Five of the 19 indices reviewed reported a form of reliability, with test–retest reliability being most commonly cited. The REALM, REALM-S, and REALM-Teen demonstrate high test–retest reliability (reliability coefficients: 0.98, 0.99, and 0.98, respectively) when administered twice, 1 week apart (Table 2). The REALM was administered to 38 prison inmates and 26 substance abuse halfway house residents [8]. The REALM-S was administered to patients in primary care clinics and REALM-Teen to adolescents from high schools, primary care clinics, and summer programs in the United States, although numbers of individuals were not specified [24,25].

Test–retest reliability was also high for the SAHLSA (Pearson correlation coefficient: 0.86). It was administered twice to 40 Spanish-speaking patients in outpatient clinics, 2 weeks apart [15,21]. The test–retest reliability of the eHEALS was assessed by administering it to adolescents and young adults in the control arm of a combined randomized trial evaluation of a eHealth literacy promotion intervention and a Web-based smoking cessation program at four time points: pre- and postintervention (time points not specified) and at 3- and 6-month follow up [18]. Test–retest reliability was calculated between scores at each interval showing modest stability from baseline to 6 months ($r = 0.60–0.40$). Reported intraclass correlation coefficient of 0.49 suggests poor stability over time [18].

There have been no published studies of intrarater reliability for indices requiring interviewer administration, and there is only one study of interrater reliability [8]. The REALM was shown to have high interrater reliability (Pearson correlation coefficient: 0.99) when administered by five research assistants to 20 patients at a university clinic [8].

Internal consistency is commonly measured by coefficient alpha and is another method for determining the degree to which items in a scale consistently measure a single construct [42]. Of the nine indices that reported internal consistency, the numeracy domain of the S-TOFHLA ($\alpha = 0.68$) and the critical health literacy domain of the FCCHL ($\alpha = 0.65$) had coefficients lower than the commonly accepted minimum $\alpha < 0.7$ [54] (Table 2).

3.2.5. Responsiveness

Responsiveness is defined as an instrument's ability to detect a clinically important change, if present, within individuals over time [55]. No studies were found that measured the responsiveness of any health literacy instrument.

3.2.6. Feasibility

This relates to how an instrument is administered and the requirements associated with the delivery. All indices that directly test individual abilities require an interviewer to administer all or some aspects of the test, although they appear easy to perform. Some interviewer training is required, particularly for the administration of the REALM and derivatives, SAHLSA and MART, which rely on subjective assessment of correct pronunciation. The estimated time to administer instruments was generally not detailed for self-reported and population measures (Tables A1 and A2 [available on the journal's Web site at www.jclinepi.com]). Of those that were administered (Table 1), the 22 minutes to complete the TOFHLA could be considered a limitation.

3.2.7. Generalizability

This refers to whether the instrument can be effectively applied in different populations and settings. The REALM-S, TOFHLA, and S-TOFHLA have been used extensively in clinical populations in the United States. They have also been administered to clinical and general populations in the United Kingdom [46,56–59], Australia [44,47], and Brazil [60] (Table 2). Investigators have found it necessary to tailor the content of the TOFHLA to local settings in the United Kingdom and Australia but not the REALM or NVS [47,57].

4. Discussion

We have critically appraised the development and psychometric properties of 19 health literacy indices. Three approaches for measuring health literacy were identified: direct testing of an individual's abilities, self-report of abilities, and population-based proxy measures. The underlying constructs assessed varied widely across instruments and most were not based on a specific conceptual framework. The content focused primarily on reading, comprehension, and numeracy skills; scoring categories were poorly defined and may not be mutually exclusive; only five had been assessed for reliability. Responsiveness to change has not been assessed for any of the indices. Overall, the TOFHLA and REALM-S had the strongest psychometric properties.

The variability in the measurement approach across instruments reflects the view that health literacy is a complex and multifaceted concept that is still evolving [61,62]. Health literacy has developed through a convergence of two main areas of study: (1) health education and promotion where health literacy is considered a personal "asset" and (2) clinical care where it is considered a "risk" that needs to be managed to achieve positive health outcomes [40,63]. Such context may account for differences in measurement approaches. For example, indices, such as the FCCHL and eHEALS, focus more on personal abilities,

whereas the TOFHLA and SBSQ look to identify individuals at risk.

However, the lack of an explicit definition of the concept that many health literacy indices were developed to measure limited our ability to make fully informed judgments about their face and content validities. Based on a widely accepted definition of health literacy [40,64], none of the indices we evaluated seemed to fully measure a person's ability to seek, understand, and use health information. This calls into question their overall relevance and usefulness.

To tailor health information to a patient's needs, a health literacy instrument should ideally be capable of clearly describing an individual patient's abilities and areas that need to be addressed. Without a clear understanding of what different categories within a health literacy measure mean, it is difficult to know how to apply this information in a clinical or public health context. At a population level, inability to clearly discriminate between groups of people with different levels of health literacy may hamper efforts to target interventions toward ''high-risk'' groups.

In the absence of a ''gold standard,'' empirical evidence of construct validity of most health literacy measures was obtained by comparison with standardized literacy tests and/or other health literacy indices. Several health literacy instruments were developed from standardized literacy tests, and therefore, it is not surprising that there was a strong correlation between them. However, there was not always a strong correlation between health literacy measures, implying that they might have measured different constructs. None of the instruments we assessed have been compared with broader elements of health literacy implied in definitions, such as communication and self-management skills [65,66].

Other authors have also identified a mismatch between definitions and measurement of health literacy [28,61,62,67] and the need for more comprehensive measures [64,68]. In a qualitative study to conceptualize health literacy from the patient perspective, we identified a range of individual abilities and contextual factors important in seeking, understanding, and using health information [69]. Individual abilities included knowing when to seek health information, verbal communication, assertiveness, capacity to process and retain information, and applying information to lifestyle, which all reflect broader constructs not incorporated in existing instruments.

Although there were no studies that specifically assessed the responsiveness of health literacy instruments, one study was identified that examined changes in health literacy over time using the REALM-S [70]. In a randomized control trial that tested the effectiveness of a literacy learning program for individuals with both depression and low health literacy, health literacy scores in the intervention group increased by an average of 7 points between baseline and final follow-up. This suggests that the REALM-S is responsive to change; however, empirical demonstration of responsiveness over time is critical if instruments are to accurately measure the effectiveness of future programs to improve health literacy in populations.

A strength of our review is the use of a structured methodological framework to critically appraise existing health literacy measures. Based on our evaluation, the REALM and TOFHLA appear to have the strongest psychometric properties, although each has its limitation. Both state a clear purpose, have been compared with a range of indices to demonstrate construct validity, are simple to understand and administer, and have been used in clinical and population settings. A limitation of our study was our inability to fully appraise the population measures because of limited information being available in the public domain.

Our study indicates that further research is needed to address psychometric gaps. For example, in-depth qualitative studies may assist in developing cutoff scores or categories that can be meaningfully applied to clinical or population settings [71]. Additionally, randomized control trials of program interventions to improve health literacy would be valuable forums to obtain information on responsiveness and reliability of existing measures. Future instruments need to encompass broader concepts of health literacy identified in existing literature [72]. This will require in-depth consultation with key stakeholders to clearly articulate a conceptual basis of health literacy and identify clear constructs to measure.

Acknowledgments

Richard H. Osborne is supported in part by a National Health and Medical Research Council (NHMRC) Population Health Fellowship (Career Development Award), and Rachelle Buchbinder is supported in part by an NHMRC Practitioner Fellowship. Joanne E. Jordan is supported by an Australian Postgraduate Award.

Appendix

Supplementary material

Supplementary material can be found, in the online version, at 10.1016/j.jclinepi.2010.04.005

References

[1] Eysenbach G, Kohler C. How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. BMJ 2002;324:573–7.

[2] Rees CE, Ford JE, Sheard CE. Evaluating the reliability of DISCERN: a tool for assessing the quality of written patient information on treatment choices. Patient Educ Couns 2002;47:273–5.

[3] Simonds SK. Health education as social policy. Health Educ Monogr 1974;2:1−25.
[4] Nutbeam D. Health promotion glossary. Health Promot Int 1998;13: 349−64.
[5] Weiss BD, Hart G, McGee DL, D'Estelle S. Health status of illiterate adults: relation between literacy and health status among persons with low literacy skills. J Am Board Fam Pract 1992;5:257−64.
[6] Dewalt DA, Berkman ND, Sheridan S, Lohr KN, Pignone MP. Literacy and health outcomes: a systematic review of the literature. J Gen Intern Med 2004;19:1228−39.
[7] USA National Academy on an Aging Society. Low health literacy skills Increase Annual Health Care Expenditures by $73 Billion. The Center for Health Care Strategies and National Academy on an Aging Society; 1999. Available at: http://www.civicengagement.org/agingsociety/publications/fact/fact_low.html. Accessed August 2009.
[8] Davis TC, Crouch M, Wills G, Abdehou D. Rapid assessment of literacy levels of adult primary care patients. Fam Med 1991;23: 433−55.
[9] Parker RM, Baker DW, Williams MV, Nurss JR. The test of functional health literacy in adults: a new instrument for measuring patients' literacy skills. J Gen Intern Med 1995;10:537−41.
[10] Del Greco L, Walop W, McCarthy R. Questionnaire development: validity and reliability. CMAJ 1987;136:699−700.
[11] Nunnally J. Psychometric theory. 2nd ed. New York: McGrath Hill; 1978.
[12] Nunally JC, Bernstein IH. Psychometric theory. New York: McGraw Hill; 1994.
[13] Jolles B, Buchbinder R, Beaton D. A study compared nine patient-specific indices for musculoskeletal disorders. J Clin Epidemiol 2005;58:791−801.
[14] Hanson-Divers EC. Developing a medical achievement reading test to evaluate patient literacy skills: a preliminary study. J Health Care Poor Underserved 1997;8:56−69.
[15] Lee SY, Bender DE, Ruiz RE, Cho YI. Development of an easy-to-use Spanish Health Literacy test. Health Serv Res 2006;41: 1392−412.
[16] Weiss BD, Mays MZ, Martz W, Casto KM, DeWalt DA, Pignone MP, et al. Quick assessment of literacy in primary care: the Newest Vital Sign. Ann Fam Med 2005;3:514−22.
[17] Chew LD, Bradley KA, Boyko EJ. Brief questions to identify patients with inadequate health literacy. Fam Med 2004;36:588−94.
[18] Norman C, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006;8:e27.
[19] Ishikawa H, Takeuchi T, Yano E. Measuring functional, communicative, and critical health literacy among diabetic patients. Diabetes Care 2008;31:874−9.
[20] Hanchate AD, Ash AS, Gazmararian JA, Wolf MS, Paasche-Orlow MK. The Demographic Assessment for Health Literacy (DAHL): a new tool for estimating associations between health literacy and outcomes in national surveys. J Gen Intern Med 2008;23:1561−6.
[21] Kutner M, Greenberg E, Jin Y, Paulsen C. The health literacy of America's adults: results from the 2003 National Assessment of Adult Literacy (NCES 2006-483). Washington, DC: National Center for Education Statistics; 2006.
[22] Rudd RE. Health literacy skills of U.S. adults. Am J Health Behav 2007;31(Suppl 1):S8−18.
[23] Canadian Council on Learning. Health literacy in Canada: initial results from the International Adult Literacy and Skills Survey 2007. Ottawa, Canada: Canadian Council on Learning; 2007.
[24] Davis TC, Long SW, Jackson RH, Mayeaux EJ, George RB, Murphy PW, et al. Rapid estimate of adult literacy in medicine: a shortened screening instrument. Fam Med 1993;25:391−5.
[25] Davis TC, Wolf MS, Arnold CL, Byrd RS, Long SW, Springer T, et al. Development and validation of the Rapid Estimate of Adolescent Literacy in Medicine (REALM-Teen): a tool to screen adolescents for below-grade reading in health care settings. Pediatrics 2006;118: e1707−14.
[26] Bass PF 3rd, Wilson JF, Griffith CH. A shortened instrument for literacy screening. J Gen Intern Med 2003;18:1036−8.
[27] Arozullah AM, Yarnold PR, Bennett CL, Soltysik RC, Wolf MS, Ferreira RM, et al. Development and validation of a short-form, rapid estimate of adult literacy in medicine. Med Care 2007;45:1026−33.
[28] Morris NS, MacLean CD, Chew LD, Littenberg B. The Single Item Literacy Screener: evaluation of a brief instrument to identify limited reading ability. BMC Fam Pract 2006;7:21.
[29] Baron-Epel O, Balin L, Daniely Z, Eidelman S. Validation of a Hebrew health literacy test. Patient Educ Couns 2007;67:235−9.
[30] Baker DW, Williams MV, Parker RM, Gazmararian JA, Nurss J. Development of a brief test to measure functional health literacy. Patient Educ Couns 1999;38:33−42.
[31] Tang Y, Pang SMC, Chan MF, Yeung GSP, Yeung VTF. Health literacy, complication awareness and diabetic control in patients with type 2 diabetes mellitus. J Adv Nurs 2007;62:74−83.
[32] Mancuso JM. Assessment and measurement of health literacy: an integrative review of the literature. Nurs Health Sci 2009;11:77−89.
[33] Taylor WS. Cloze procedure: a new test for measuring readability. Journalism Q 1953;30:415−33.
[34] Williams MV, Parker RM, Baker DW, Parikh NS, Pitkin K, Coates WC, et al. Inadequate functional health literacy among patients at two public hospitals. JAMA 1995;274:1677−82.
[35] Baker DW, Parker RM, Williams MV, Pitkin K, Parikh NS, Coates W, et al. The health care experience of patients with low literacy. Arch Fam Med 1996;5:329−34.
[36] Ishikawa H, Nomura K, Sato M, Yano E. Developing a measure of communicative and critical health literacy: a pilot study of Japanese office workers. Health Promot Internation 2008;23:269−74.
[37] Nutbeam D. Health literacy as a public health goal: a challenge for contemporary health education and communication strategies into the 21st century. Health Promot Internation 2001;15:259−67.
[38] Norman CD, Skinner HA. eHealth Literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006;8:e9.
[39] Ratzan SC, Parker RM. Health literacy—identification and response. J Health Commun 2006;11:713−5.
[40] Nutbeam D. The evolving concept of health literacy. Soc Sci Med 2008;67:2072−8.
[41] Australian Bureau of Statistics. Health literacy, Australia 2006. Canberra, Australia: Australian Bureau of Statistics; 2008.
[42] Portney LG, Watkins MP. Foundations of clinical research—applications to practice. Stamford, CT: Appleton & Lange; 1993.
[43] Child D. Essentials of factor analysis. 2nd ed. London, UK: Cassell; 1990.
[44] Buchbinder R, Hall S, Youd JM. Functional health literacy of patients with rheumatoid arthritis attending a community-based rheumatology practice. J Rheumatol 2006;33:879−86.
[45] Chisolm DJ, Buchanan L. Measuring adolescent functional health literacy: a pilot validation of the Test of Functional Health Literacy in Adults. J Adolesc Health 2007;41:312−4.
[46] Ibrahim SY, Reid F, Shaw A, Rowlands G, Gomez GB, Chesnokov M, et al. Validation of a Health Literacy Screening Tool (REALM) in a UK population with coronary heart disease. J Public Health (Oxford) 2008;30:449−55.
[47] Barber M, Osborne R, Staples M, Clerehan R, Elder C, Buchbinder R. Up to a quarter of the Australian population may have suboptimal health literacy depending upon the measurement tool: results from a population-based survey. Health Promot int 2009;24: 252−61.
[48] Osborn CY, Weiss BD, Davis TC, Skripkauskas S, Rodrigue C, Bass PF, et al. Measuring adult literacy in health care: performance of the Newest Vital Sign. Am J Health Behav 2007;31(Suppl 1):S36−46.
[49] Mossenson L, Hill P, Masters G, et al. TORCH: Test of Reading Comprehension Manual. Melbourne: Australian Council for Educational Research; 1998.

[50] Chew L, Griffin JM, Partin MR, Noorbaloochi S, Grill JP, Snyder A, et al. Validation of screening questions for limited health literacy in a large VA outpatient population. J Gen Intern Med 2008;23:561−6.
[51] Hu L, Bentler P. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Modeling 1999;6:1−55.
[52] Bombard JM, Powell KE, Martin LM, Helmick CG, Wilson WH. Validity and reliability of self-reported arthritis: Georgia senior centers, 2000-2001. Am J Prev Med 2005;28:251−8.
[53] Reijman M, Hazes JM, Koes BW, Verhagen AP, Bierma-Zeinstra SM. Validity, reliability, and applicability of seven definitions of hip osteoarthritis used in epidemiological studies: a systematic appraisal. Ann Rheum Dis 2004;63:226−32.
[54] Hays R, Revicki D. Reliability and validity (including responsiveness). In: Fayers P, Hays RD, editors. Assessing quality of life in clinical trials. Oxford, UK: Oxford University Press; 2005. p. 25−39.
[55] Cardol M, Beelen A, van den Bos GA, de Jong BA, de Groot IJ, de Haan RJ. Responsiveness of the impact on participation and autonomy questionnaire. Arch Phys Med Rehabil 2002;83:1524−9.
[56] Gordon MM, Hampson R, Capell HA, Madhok R. Illiteracy in rheumatoid arthritis patients as determined by the Rapid Estimate of Adult Literacy in Medicine (REALM) score. Rheum (Oxford) 2002;41:750−4.
[57] von Wagner C, Knight K, Steptoe A, Wardle J. Functional health literacy and health-promoting behaviour in a national sample of British adults. J Epi Community Health 2007;61:1086−90.
[58] Rutherford J, Holman R, MacDonald J, Taylor A, Jarrett D, Bigrigg A. Low literacy: a hidden problem in family planning clinics. J Fam Plann Reprod Health Care 2006;32:235−40.
[59] Dani KA, Stobo DB, Capell HA, Madhok R. Audit of literacy of medical patients in north Glasgow. Scott Med J 2007;52:21−4.
[60] Carthery, -Goulart MT, Anghinah R, Areza-Fegyveres R, Bahia VS, Brucki SM, Damin A, et al. Performance of a Brazilian population on the test of functional health literacy in adults. Revista de Saude Publica 2009;43:631−8.
[61] Baker DW. The meaning and the measure of health literacy. J Gen Intern Med 2006;21:878−83.
[62] Dray S, Papen U. Literacy and health: towards a methodology for investigating patients' participation in healthcare. J Appl Linguist 2004;1.3:311−32.
[63] Zarcadoolas C, Pleasant A, Greer DS. Understanding health literacy: an expanded model. Health Promot Int 2005;20:195−203.
[64] Nielsen-Bohlman L, Panzer A, Kindig D. Health literacy: a prescription to end the confusion. Washington, DC: National Academies Press; 2004.
[65] Osborne RH, Elsworth GR, Whitfield K. The Health Education Impact Questionnaire (heiQ): an outcomes and evaluation measure for patient education and self-management interventions for people with chronic conditions. Patient Educ Couns 2007;66:192−201.
[66] Cockburn J, Fahey P, Sanson-Fisher R. Construction and validation of a questionnaire to measure the health beliefs of general practice patients. Fam Pract 1987;4:108−16.
[67] Entwistle V, Williams B. Health literacy: the need to consider images as well as words. Health Expect 2008;11:99−101.
[68] Sihota S, Lennerd L. Health literacy—being able to make the most of health. London, UK: National Consumer Council UK; 2004.
[69] Jordan JE, Briggs AM, Brand CA, Osborne RH. Enhancing patient engagement in chronic disease self-management support initiatives in Australia: the need for an integrated approach. Med J Aust 2008;189:S9−S13.
[70] Weiss BD, Francis L, Senf JH, Heist K, Hargraves R. Literacy education as treatment for depression in patients with limited literacy and depression: a randomized controlled trial. J Gen Intern Med 2006;21:823−8.
[71] Dew K. A health researcher's guide to qualitative methodologies. Aust N Z J Public Health 2007;31:433−7.
[72] Jordan JE, Buchbinder R, Osborne RH. Conceptualising health literacy from the patient perspective. Patient Educ Couns 2010;79:36−42.